Subjective probability distributions, a psychometric approach.

Terlouw, Pieter

SUMMARY

A motto in a number of disciplines seems to be 'Measure it several times (with different devices). Take some sort of a mean. That mean, in general , will be a better estimate, than the estimate obtained with a single measurement'. It is also desirable in several disciplines to know how reliable these estimates are. These aspects combined with the validity question form the main subjects of this thesis. The thesis is restricted to subjective probability distributions for proportion questions.

The following questions for elicited subjective probability distributions were considered: 'How reliable are they?', 'How valid are they?', and 'Does any tool exist to improve reliability and/or validity?'. Beside these questions, attention was paid to the fit of members of suitable parametric families of subjective probability distributions to elicited distributions.

RELIABILITY

After an introduction in chapter 1, in which the aims and questions of the research reported were given, in chapter 2 the required notions were discussed, concerning the area of subjective probability distributions, to be used in the next chapters. In chapter 3, based on a classical test theory model ('score is true score plus error', X = T + E), a similar model was proposed and developed for subjective probability distributions. The triple model:

$$(M,W,S) = (\tau_m,\tau_w,\tau_s) + (E_m,E_w,E_s)$$

in which m is a measure for the central tendency, w a measure for the width, and s a measure for the skewness of a subjective probability distribution, formed, combined with five assumptions, the basis for the definition of reliability of subjective probability distributions. Reliability of m, w, and s could be calculated as a simple correlation

for repeated measurements. Reliability was defined per person as well as per question.

In the literature remarkably little attention was paid to the reliability of measures for central tendency, while even less attention was given to the reliability of the width. It should be noted, however, that width is an essential characteristic of subjective probability distributions. It is also a necessary characteristic to be used in several calculations, for instance, in a Bayesian analysis. Moreover, if reliability calculations were made, they were almost always made at group level.

VALIDITY

Through concepts like overconfidence, in the literature much attention was given to validity. Whether or not an elicited subjective probability distribution is a valid representation of knowledge is hard to establish. In chapter 3 different approaches for such an establishment were considered. The most promising one seemded to be: construct validity, an ongoing process of checking predictions with respect to the elicited subjective probability distribution.

FITTED SUBJECTIVE PROBABILITY DISTRIBUTIONS

In a Bayesian analysis an elicited subjective probability distribution is used as a prior distribution. For calculation purposes one has to know the formula of the prior distribution used. Elicitation methods seldomly lead to such a formula. Therefore, one has to fit a member of a family of distributions to elicited characteristics of a subjective probability distribution.

In chapter 4 four possible families of distributions to be fitted were considered: 1. beta distributions; 2. triangular distributions; 3. a mixture of beta distributions consisting of a fitted beta distribution combined, for a small part, with a flat beta distribution (an insurance for overconfidence); and 4. a mixture of triangular distributions consisting of a fitted triangular distribution combined, for a small part, with a triangular distribution on

(0,1) of which the mode equals the elicited mode. In chapter 4 two methods of fit were discussed: 1. using losses (distance measures between elicited and fitted distribution) aiming at minimal distance; and 2. using approximation formulas from which the parameters of the distribution to be fitted follow immediately. In chapter 4 some attention also was given to two types of elicitation methods: direct and indirect. It was argued that for proportion questions a direct method was more suitable.

The computer program SPAT, an elicitation instrument for subjective probability distributions for proportion questions was discussed in chapter 5. The general ideas given in the first paragraph of this summary were the main guidelines for the construction of SPAT. Lourens (1984) already showed that the use of SPAT led to a reduction of the number of mistakes in the specification of a subjective probability distribution. Overconfidence, nevertheless, remained.

EXPERIMENTAL RESULTS

The triple model proved to be a useful model. Reliabilities, therefore, could be determined. In two different experiments it was shown that the reliability of the mode per person as well as per question was high. Width of a subjective probability distribution was fairly reliable per person as well as per question. The reliability of the skewness per person as well as per question was low. That low reliability limited the validity of the skewness as well.

In fitting members of families of distributions to elicited characteristics there was no remarkable difference between members of the family of beta distributions and members of the family of triangular distributions. Taking into account the advantages of the use of a beta distribution (it is defined on (0,1), and it is mathematically elegant in calculations) one could give preference to the use of beta distributions as a family of distributions to be fitted to elicited characteristics of subjective probability distributions. With respect to the fit of mixtures of distributions a surprising result appeared: a 50%-insurance, to obtain an optimal

score on a scoring rule. Further research is needed on the use of mixtures of distributions (in elicitation methods).

Concluding, subjects specify mode and width of a subjective probability distribution (fairly) reliably. Third, and higher order moments are specified unreliably. There is no motive for using another distribution than a beta distribution to be fitted to elicited characteristics of a subjective probability distribution.